

A Support System for Visually Impaired Persons to Understand Three-dimensional Visual Information Using Acoustic Interface

Yoshihiro Kawai

Fumiaki Tomita

National Institute of Advanced Industrial Science and Technology (AIST)

Tsukuba Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

y.kawai@aist.go.jp

f.tomita@aist.go.jp

Abstract

Visual information processing technology is very important in the implementation for sensory substitution of visually impaired persons as well as applications to factory automation. This paper outlines the design of a visual support system that provides 3D visual information using 3D virtual sounds. Three-dimensional information, such as distance map, object recognition, and object tracking required for the visually impaired user, is obtained by analyzing images captured by stereo cameras. Using a 3D virtual acoustic display, which relies on Head Related Transfer Functions (HRTFs), the user is informed of the locations and movements of objects. The user's external auditory sense is not impeded as the system uses bone conduction headphones which do not block out environment sounds. The proposed system is expected to be useful in the situations where the infrastructure is incomplete and the situation changes in real-time. We plan experiments using this system to guide users while walking and playing sports.

1 Introduction

Much of the information that humans acquire from the outside world is obtained through sight. Without this facility visually impaired people suffer inconveniences in their daily and social lives. Therefore, much research has been done worldwide on support systems for the visually impaired [5, 6, 9]. However, there are still many problems in representing the real-time information that is changing around the user.

Long canes and seeing-eye dogs are widely used as walking support devices in active operator support systems. However, the range that a user can sense with the canes is limited, and the use of seeing-eye dogs, still entails problems of availability and practicability. Support devices using electronic technologies have been developed, but considerable training is needed to use them. It is important to prepare an infrastructure that enables users to easily under-

stand the circumstances in their periphery. Surface bumps, braille panels, and audio traffic signals for visually impaired persons are in use. However, economic realities limit their installation and availability. Many problems cannot be solved only by infrastructure maintenance and development. Therefore, we aimed to develop an active support system - an intelligent support device - that would provide access to 3D spatial information surrounding the user by using an acoustic interface.

Among other reported visual aid systems using sound are a system that uses ultrasonic waves, one that displays images by differences of frequency pitch and loudness [2, 7] and one that utilizes stereophonic effects. However, the target of these systems is mainly a 2D space. Three-dimensional sound can provide more real-world information because it includes an intuitive feeling of depth and a feeling of front and rear.

We are developing a support system that displays 3D visual information using 3D virtual sound. It is unique in that 3D environment information is acquired for the task that the user sets, and it is represented by 3D virtual sound. Images captured by small stereo cameras are analyzed in the context of a given task to obtain the 3D structure, and object recognition is performed. The results are then conveyed to the user via 3D virtual sound. This system is expected to be useful in situations where the infrastructure is incomplete and the situation surrounding the user changes in real-time. In addition, this system can be used without much learning because it provides information via virtual sound superposed on the actual environment sounds. This method would not replace or impede the user's existing external auditory sense. We assume that it could be used, for example, to assist while walking or playing sports. The importance to walking assistance needs no further explanation, however, assistance while playing sports is certainly also important. The sports that visually impaired persons can do are limited to those with 2D movement, such as jogging, roller skating, floor volleyball, etc. There are many requests voiced by those who wish to play sports with 3D movements to-



Figure 1. Support system overview.

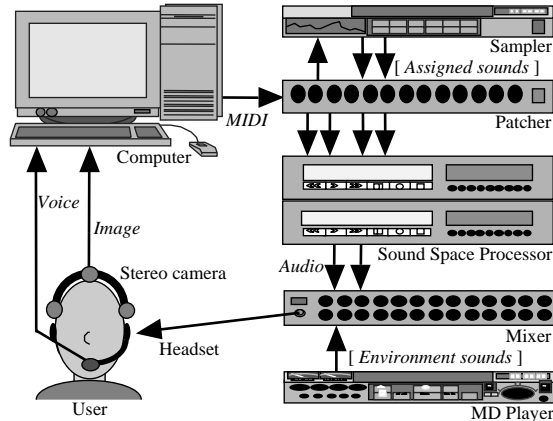


Figure 2. System composition.

gether with sighted people. So we believe that our proposed system is one answer that satisfies their needs.

2 System hardware

We have built the prototype system shown in Figure 1 to develop a visual support system to perform experiments on visual information processing, device control, and sound expression. It has a stereo camera system with three small cameras, and a headset with a microphone and headphones. Figure 2 shows the configuration of this system.

2.1 Stereo camera system

We use small cameras to acquire information on objects and visual environments. The captured images are analyzed to obtain 3D information. The advantages of this method are that it is suited for object recognition and tracking, measurement of distant objects (e.g., discrimination of the red/green light of a traffic signal from far away) and for character information readability (e.g., characters on a sign). Although it is still difficult to analyze images to obtain 3D visual information, there exists a potential use for recent pattern recognition techniques in our application. Our developing vision system enables analysis of stereo images captured by three cameras, reconstruction of 3D objects in the target scene, recognition by model matching,

and tracking of moving objects.

It is desirable for the visual information input unit to be small and light weight because these devices will be mounted on the user's head. However, high performance is required for accurate measurements. As a result, we mounted an aluminum frame with three small cameras on a helmet. We have set the focus of the lens at more than 2 m, a point at which a user's cane cannot reach.

2.2 Acoustic system

Recently, with the development of virtual reality technologies, the technical progress of acoustics in a virtual space has been remarkable. We can use 3D virtual sound easily, since some 3D sound equipment has already been produced and is available commercially. We have built our acoustic system around the RSS-10 sound space processor made by Roland corporation (shown on left in Figure 1). This device enables us to calculate an arbitrary 3D virtual sound space on the basis of HRTFs[1] by just input position, movement vector, and sound source. For the output device, we selected bone conduction headphones, which do not entirely cover the user's ears, and therefore do not impede hearing or understanding environment sounds. It is very important not to degrade the user's external auditory sense.

2.3 System control

We will explain the design of the whole system, as shown in Figure 2. The three images captured by the stereo cameras are sent to a computer and analyzed by the process mentioned above. After the 3D data is reconstructed, the user's targets are recognized and tracked. Both the status of the objects acquired and the sound assigned to each object from a sampler are input to the sound space processor. A sound image for each target is mapped in a 3D virtual sound space and transmitted to the user. Although as yet unfinished, we may also have a microphone for the user to set a task by voice, and this may be processed by a voice recognition engine.

3 Three-dimensional visual information processing method

We briefly explained that the method of obtaining 3D information would use three captured images. The features of our proposed method is although only range information and color information is obtained, 3D shape recognition and tracking are possible. We will also show the results of a 3D visual information processing experiment on playing catch ball as one application example, which is an activity in which many visually impaired people wish to participate.

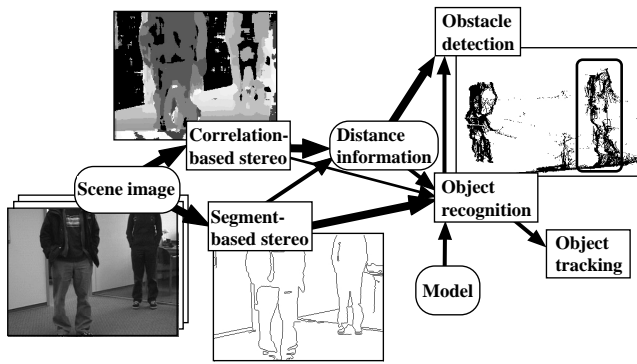


Figure 3. Flowchart of 3-D vision algorithm.

3.1 Processing methods

Measurement, recognition, and tracking of 3D objects in the target scene is done by analysis of stereo images. The process flow is shown in Figure 3, and it is an integration of two stereo algorithms, a segment-based method and a correlation-based method. Recognition and tracking processes are performed based on the results of the segment-based method.

The segment-based method, a structural analysis method, is an algorithm for reconstructing 3D wire-frames by correspondence of boundary edges [4]. First, some special features, such as segments, are extracted after detecting boundary edges, and a correspondence search for them is performed. While is a complicated procedure, it is a superior method for structure reconstruction and recognition of 3D objects.

On the other hand, distance information obtained by the correlation-based method, calculates the disparities between stereo images using the fact that correlation values of intensity at the same place are higher. A disadvantage is that it requires a long processing time if the search range is not limited, however, with its simple algorithm, it can be processed in real-time using special hardware. The 3D data obtained comprise sets of points, and a structuring process such as segmentation is needed.

After acquiring the 3D data in stereo vision, the object recognition process follows [8]. The 3D data are matched with object models in a database to identify what objects are present and to determine their status.

In 3D Euclid space, a 4×4 transformation matrix T is defined with a 3×3 rotation matrix and a translation vector. In other words, recognition means the calculation of T by comparing the models with the data. This process has two phases: an initial matching phase and a fine adjustment phase. In the initial matching phase, T is roughly calculated by comparing the geometric features of the model with the data. Then, accuracy is iteratively improved using the whole data for candidates, which are higher than the

threshold. The best result is selected to correspond with the mode.

Through this process, users can know information on an object required to perform a task. In addition, gaps or obstacles are detected using depth information obtained by the correlation-based method.

The tracking process is performed when it is necessary to track a moving object. The recognition result is used as an initial position. There are two algorithms for tracking an object. One is using a algorithm similar to the latter phase in the recognition process where T is updating each time. And the other uses features, a comparatively light process, such as color segmentation. These two methods are distributed according to the situation. If the boundary edges of the target objects in the images are sufficiently obtained, the first tracking method is applied. Otherwise, the other is used. Accurate position and posture of objects can be obtained using the former method, but the processing performance is less than 10 fps without a special hardware device. The details of this algorithm have been reported [3]. The feature method is applied, for example, when the object has special features, such as characteristic colors, and can be easily divided from the background using segmentation in the RGB color space. However, the accuracy of position detection is lower than with the former algorithm and the posture is not calculated. The sample experiment described in section 3.2, the tracking method is based on the feature method, because the boundary edges of the object cannot be sufficiently detected because of the small image size (160×120 pixel). For this ball tracking example, there is no problem using this algorithm.

Finally, the results of these measurement, recognition and tracking processes are transmitted to the virtual sound system described in section 2.2 to represent 3D visual information using sounds.

3.2 Experiment of ball tracking

Technology for tracking a moving object is necessary to help visually impaired persons play sports, for example, playing catch ball. The task involves detection of a moving ball. We will describe an experiment result on vision process for playing catch ball.

Figure 4 (a) shows a captured stereo image using the stereo camera system in Figure 1, which is mounted on the helmet. The position and direction of the stereo cameras, where the rows and columns of the cameras correspond to epipolar lines, are laid out in order to reduce searching time with epipolar constraints. In this experiment, we performed a test of tracking a sponge ball. Its color is red, blue, and yellow. First, a pitcher held a beanbag in both hands, and then throw it to the system. The number of total images was 123 frames per 8.5 sec. (14.5 fps). The image size was 160×120 pixels, but at first, some images were captured at

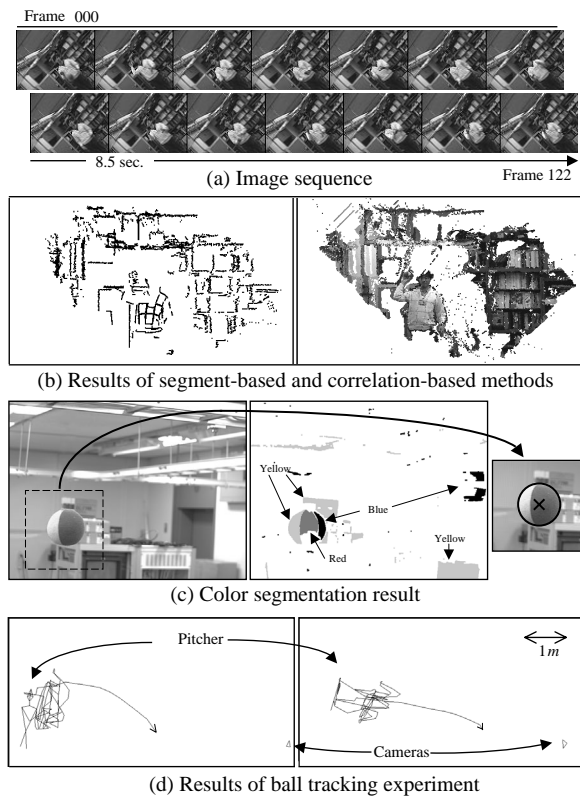


Figure 4. Results of thrown ball tracking experiment.

640×480 pixels (4.5 fps) to detect the position of the target ball. The pitcher had to show the ball clearly to enable it to be recognized easily (to calculate initial parameters).

After recognizing it, the positions were tracked using the algorithm in section 3.1. The reconstructed result for one scene (frame number 099) using the segment-based method is shown in the left figure of Figure 4 (b). Distance measurement using the correlation-based method is also done at the same time to detect obstacle objects (the right figure in Figure 4 (b)). The tracking process using only the color segmentation method described in section 3.1 was performed in this experiment to track a fast-moving object, the thrown ball. The ball included primary colors (red, blue, and yellow), which are easily segmented in the RGB color space (See Figure 4 (c)). In each camera's image, the regions with each of the color parts are segmented, and only one region, the region boundary of each color space, are selected. For this region, a circle was fitted and the 2-D coordinates (col , row) of the center were calculated. Then, the position of the moving ball (x , y , z) was calculated using three positions (col_{Left} , row_{Left}), (col_{Center} , row_{Center}), (col_{Right} , row_{Right}). This process was iterated repeated to trace the ball. The time (frame number i) requires to search the new position was reduced by guessing it from obtained transformation parameters T_i ($i=0..n-1$). The re-

sult of tracking the ball is shown in Figure 4 (d). Although the position resolution was low because small images were used, the tracked path was fairly correct.

Small cameras with fixed focus and mounted on a frame are used as the input device. They do not comprise active camera system, so there is a problem in that the object information is not reconstructed closer than the 2 m distance (See Figure 4 (d)), because it was not captured in every camera. However, this camera setting was decided based on the use of a long cane, as was described in section 2.1. Also, this experiment was carried out off-line, so the computation time was not in real time. We are now developing a real-time on-line system.

4 Conclusions

We are developing a recognition support system using 3D virtual sound to enable visually impaired persons to obtain reliable information about the environment. The design of our prototype system, 3D visual and auditory information processing methods, and some basic experimental results on our acoustic tests are described. We showed that it is a usable system if tasks are limited, though there is still a problem of processing time for image processing. Problems in sound image localization were clarified through the experiments.

In the future, we will first complete this as an on-line system and develop algorithms for computer vision to analyze 3D visual information.

References

- [1] J. Blauert, "Spatial Hearing (Revised Edition)", The MIT Press, 1996.
- [2] T. Ifukube, T. Sasaki, C. Peng: "A blind mobility aid modeled after echolocation of bats", IEEE Trans. BME-38, 5, pp.461-465, 1991.
- [3] Y. Ishiyama, Y. Sumi, F. Tomita, "3-D Motion Tracking of 3-D Objects Using Stereo Vision", Journal of RSJ, 18,2 , pp.213-220, 2000 [in Japanese].
- [4] Y. Kawai, T. Ueshiba, Y. Ishiyama, Y. Sumi, F. Tomita, "Stereo Correspondence Using Segment Connectivity", Proc. of ICPR'98, I, pp.648-651, 1998.
- [5] K. Koshi, H. Kani, Y. Tadokoro, "Orientation Aids for the Blind Using Ultrasonic Signpost System", The 1st Joint Meeting of BMES and EMBS, pp.587, 1999.
- [6] J. M. Loomis, R. G. Golledge, R. L. Klatzky., J. M. Speige, J. Tietz, "Personal guidance system for the visually impaired", Proc. of ASSETS'94, pp.85-91, 1994.
- [7] P. B. L. Meijer, "An Experimental System for Auditory Image Representations", IEEE Trans. Biomed. Eng., 39, 2, pp.112-121, 1992.
- [8] Y. Sumi, Y. Kawai, T. Yoshimi, F. Tomita, "Recognition of 3D free-form objects using segment-based stereo vision", Proc. of ICCV'98, pp.668-674, 1998.
- [9] D. H. Warren, E. R. Stelow, "Electronic Spatial Sensing for the Blind", Martinus Nijhoff Publishers, pp.35-61, 1985.